



Building a Data Set of the Baltimore Symphony Orchestra's Performance History

Matt Testa, Peabody Institute of the Johns Hopkins University

Andre Avorio, Open Music Library



Partners and background



Chapter grant from MLA in 2017 to hire an intern, trained and supervised by staff at Peabody's Friedheim Library



Purpose is to create a data set that can be open to researchers and the public

OML to harvest the data into their performance history site

gulp

C	J	K	L	M	N	O
Date	Conductor	Soloist(s)	Composer	Work Title	Uniform	Work
1/10/1997	Pinchas Zukerman	Pinchas Zukerman, Viola	Bartók, Béla	Concerto for Viola and Orchestra, Opus Post	CONCERTO, VIOLA AND ORCHESTRA	2
1/10/1997	Pinchas Zukerman	Pinchas Zukerman, Violin	Bach, Johann Sebastian	Concerto No. 1 in A minor for Violin and Str	CONCERTO, VIOLIN, NO. 1, A MINOR, I	1
1/11/1997			Intermission,	INTERMISSION	INTERMISSION	3
1/11/1997	Pinchas Zukerman		Mendelssohn, Felix	Symphony No. 4 in A major, opus 90, ?Italiar	SYMPHONY NO. 4, A MAJOR, OPUS 90	4
1/11/1997	Pinchas Zukerman	Pinchas Zukerman, Viola	Bartók, Béla	Concerto for Viola and Orchestra, Opus Post	CONCERTO, VIOLA AND ORCHESTRA	2
1/11/1997	Pinchas Zukerman	Pinchas Zukerman, Violin	Bach, Johann Sebastian	Concerto No. 1 in A minor for Violin and Str	CONCERTO, VIOLIN, NO. 1, A MINOR, I	1
1/16/1997		David Stambler, Clarinet	Intermission,	INTERMISSION	INTERMISSION	10
1/16/1997	Erich Kunzel	David Stambler, Clarinet	Basie, Count	One O'Clock Jump	ONE O'CLOCK JUMP	4
1/16/1997	Erich Kunzel	David Stambler, Clarinet	Bassman, George	I'm Gettin' Sentimental Over You	I'M GETTIN' SENTIMENTAL OVER YOU	14
1/16/1997	Erich Kunzel	David Stambler, Clarinet	Bishop, Joe	Woodchopper's Ball	WOODCHOPPER'S BALL	8
1/16/1997	Erich Kunzel	David Stambler, Clarinet	Brown, Les/, Horner, Ben	Sentimental Journey	SENTIMENTAL JOURNEY	3
1/16/1997	Erich Kunzel	David Stambler, Clarinet	Ellington, Duke/, Tizol, Juan	Caravan	CARAVAN	5
1/16/1997	Erich Kunzel	David Stambler, Clarinet	Garland, Joseph C.	In the Mood	IN THE MOOD	2



Questions

- What kinds of research questions would we want this data set to help answer?
- What could we learn from similar published data sets of other performing arts organizations?
- What tools and techniques would be most efficient in standardizing a complex and messy data set?

Cluster & Edit column "Composer"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more ...](#)

Method

Distance Function

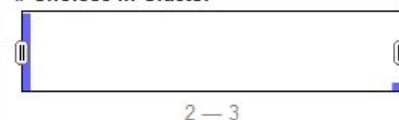
Radius

Block Chars

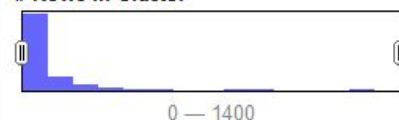
123 clusters found

3	46	<ul style="list-style-type: none"> FaurŽ, Gabriel-Urbain (20 rows) Faur., Gabriel-Urbain (19 rows) Faure, Gabriel-Urbain (7 rows) 	<input type="checkbox"/>	FaurŽ, Gabriel-Urbain
3	67	<ul style="list-style-type: none"> Rimsky-Korsakov, Nicolai (65 rows) Rimsky Korsakov, Nocolai (1 rows) Rimsky Korsakov, Nicolai (1 rows) 	<input type="checkbox"/>	Rimsky-Korsakov, Nicolai
3	7	<ul style="list-style-type: none"> Janacek, Leos (5 rows) Janřcek, Leos (1 rows) Janecek, Leos (1 rows) 	<input type="checkbox"/>	Janacek, Leos
3	70	<ul style="list-style-type: none"> Suppe, Franz von (46 rows) SuppŽ, Franz von (14 rows) Supp., Franz Von (10 rows) 	<input type="checkbox"/>	Suppe, Franz von

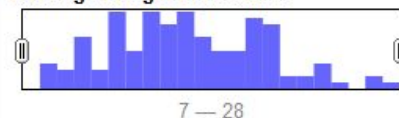
Choices in Cluster



Rows in Cluster



Average Length of Choices

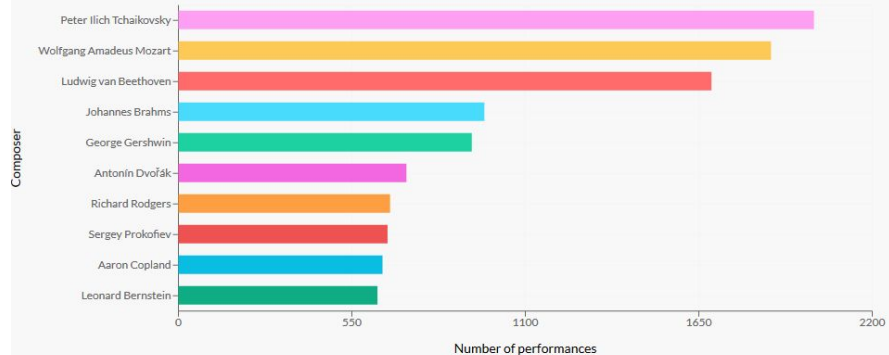


Transformation

- Visualization of data
- Links to related materials in OML on composers and works

Baltimore Symphony Orchestra

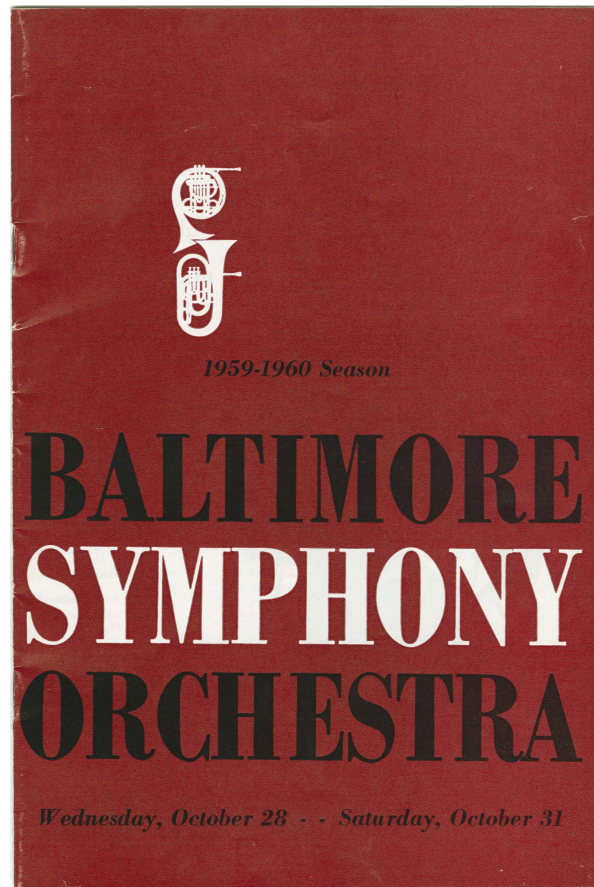
7,593 events





What's next?

- More data cleanup
- Improving search functionality
- MOAR DATAS:
(1916-1967, 2017-present)
- Next data set



Beta site:

openmusiclibrary.org/events/

raw data: <https://github.com/mtesta3/Baltimore-symphony>

Matt Testa: mtesta3@jhu.edu

Andre Avorio: AAvorio@alexanderstreet.com
